Joint Blind Source Separation of Multidimensional Components: Model and Algorithm

Dana Lahat, Christian Jutten

JOINT BLIND SOURCE SEPARATION OF MULTIDIMENSIONAL COMPONENTS: MODEL AND ALGORITHM

Dana Lahat and Christian Jutten

GIPSA-Lab, UMR CNRS 5216, Grenoble Campus, 38400 Saint Martin d'Hères, France

ABSTRACT

This paper deals with joint blind source separation (JBSS) of multidimensional components. JBSS extends classical BSS to simultaneously resolve several BSS problems by assuming statistical dependence between latent sources across mixtures. JBSS offers some significant advantages over BSS, such as identifying more than one Gaussian white stationary source within a mixture. Multidimensional BSS extends classical BSS to deal with a more general and more flexible model within each mixture: the sources can be partitioned into groups exhibiting dependence within a given group but independence between two different groups. Motivated by various applications, we present a model that is inspired by both extensions. We derive an algorithm that achieves asymptotically the minimal mean square error (MMSE) in the estimation of Gaussian multidimensional components. We demonstrate the superior performance of this model over a two-step approach, in which JBSS, which ignores the multidimensional structure, is followed by a clustering step.

***Index Terms*—** Joint BSS; independent vector analysis; multidimensional ICA; independent subspace analysis

1. INTRODUCTION

In this work, we present a model inspired by two recently-proposed extensions to blind source separation (BSS) that until now have been dealt with only separately: 1) independent vector analysis (IVA), also termed joint BSS (JBSS) [1, 2] and 2) multidimensional independent component analysis (MICA), also termed independent subspace analysis (ISA) [3, 4]. The new model, termed joint ISA (JISA), can be regarded as a generalization of JBSS to multidimensional components.

The idea of solving the dependent sources / multidimensional components problem in terms of subspace separation through ICA was first demonstrated in [5], on fetal electrocardiography (ECG) recordings. The perspective of *multidimensional* ICA, of vector-valued components whose representation is based on unambiguous projections on the sources' respective subspaces, was presented in [5], and an elaborate geometric framework was suggested in [3]. Performance bounds and identifiability results of second-order BSS of piecewise-stationary multidimensional data can be found in [6–8]. The *advantage* of using a true multidimensional model over the more prevalent two-step approach of ICA followed by a clustering step [9] has recently been quantified analytically for second-order BSS of piecewise-stationary multidimensional data [8].

Multidimensional data may occur due to various complex relations within the dependent elements. The dimension of a dependent group may not always reflect the actual number of its underlying elements. As a result, in multidimensional models, there is not always a physically meaningful interpretation to separating the multidimensional components back into single-dimensional elements. This holds, for example, in neurological activity observed by functional magnetic resonance imaging (fMRI) [10], fetal ECG [3, 5], natural images [4] and astrophysical processes [11]. Hence, a one-dimensional model for real-world data is often just an approximation.

The idea to simultaneously solve several BSS problems by exploiting higher-order statistical dependence between latent sources *across* mixtures was introduced by Kim et al. [1], and termed IVA. The method has been shown to be able to resolve the permutation ambiguity that is inherent to classical BSS up to a single permutation matrix common to all mixtures. Li et al. [2] have shown that the JBSS framework provides sufficient constraints for identifying multiple white stationary real Gaussian sources, a problem that is non-identifiable with classical BSS. It has later been shown that JBSS can be solved also via generalized matrix diagonalization that minimizes a quadratic criterion, and using either second- or higher-order statistics [12]. Recently, JBSS algorithms that minimize the maximum likelihood (ML), mutual information (MI) and entropy have been proposed [13, 14]. A comprehensive theoretical analysis of JBSS can be found in [14] and references therein. Considering the growing evidence of JBSS as a helpful tool in various applications, such as multiset data analysis [2, 12, 14] and dynamic systems [15], and the fact that natural signals are often better modelled as multidimensional, it is only natural to take advantage of the benefits of both.

In the following, †, $E\{\cdot\}$, and $\mathrm{tr}\{\cdot\}$ denote transpose, expectation and trace, respectively. Given a vector $\mathbf{b}$ of posi-

This work is funded by the project CHESS, 2012-ERC-AdG-320684. GIPSA-Lab is a partner of LabEx PERSYVAL-Lab (ANR–11-LABX-0025).

tive integers that sum up to m, $\mathrm{bdiag}_{\mathbf{b}}\{\mathbf{M}\}$ extracts from an $m \times m$ matrix $\mathbf{M}$ a block-diagonal matrix with block-pattern $\mathbf{b}$. For simplicity, we assume that all values are real.

The rest of this paper is as follows. In Sec. 2 we present and define the model that we denote JISA, and formalize it mathematically. In Sec. 3 we present a relative gradient (RG) algorithm that achieves asymptotically optimal component estimation for Gaussian data. Numerical simulations in Sec. 4 validate the correct functionality of the proposed algorithm. We conclude our work with a discussion in Sec. 5.

2. PROBLEM FORMULATION

Consider T observations of K vectors $\mathbf{x}^{[k]}(t)$, modelled as

$$\mathbf{x}^{[k]}(t) = \mathbf{A}^{[k]}\mathbf{s}^{[k]}(t) \quad 1 \le t \le T\,,\, 1 \le k \le K \qquad (1)$$

where $\mathbf{A}^{[k]}$ are invertible $m \times m$ matrices that may differ $\forall k$. Given the partition $\mathbf{s}^{[k]}(t) = [\mathbf{s}_1^{[k]\dagger}(t), \ldots, \mathbf{s}_n^{[k]\dagger}(t)]^\dagger$, where $\mathbf{s}_i^{[k]}(t)$ are $m_i \times 1$ vectors, $n \le m$, $m_i \ge 1$ and $\sum_{i=1}^n m_i = m$, the model that we define as JISA corresponds to the assumption that all the elements of the $Km_i \times 1$ vector $\mathbf{s}_i(t) = [\mathbf{s}_i^{[1]\dagger}(t), \ldots, \mathbf{s}_i^{[K]\dagger}(t)]^\dagger$ are statistically *dependent* and the pairs $(\mathbf{s}_i(t), \mathbf{s}_j(t))$ are statistically *independent* for all $i \ne j \in \{1, \ldots, n\}$. Given the pattern $\mathbf{m} = [m_1, \ldots, m_n]^\dagger$ and the set of observations $\mathcal{X} = \{\mathbf{x}^{[k]}(t)\}_{k=1,t=1}^{K,T}$, the *problem* of JISA is that of finding full-rank matrices $\mathbf{A}^{[k]}$ such that the source vectors $\mathbf{s}_1(t), \ldots, \mathbf{s}_n(t)$ are as independent as possible. This notion is given a definite meaning in Section 2.1 where we set up a simple statistical model that, via its likelihood function, yields a quantitative measure of independence. When $m_i = 1\ \forall i$, we obtain the regular JBSS.

The above partition of $\mathbf{s}^{[k]}(t)$ induces a corresponding partition in the mixing matrices: $\mathbf{A}^{[k]} = [\mathbf{A}_1^{[k]}, \ldots, \mathbf{A}_n^{[k]}]$ with $\mathbf{A}_i^{[k]}$ the ith $m \times m_i$ column-block of $\mathbf{A}^{[k]}$. The multiplicative model (1) may now be rewritten as a sum of $n \le m$ *components*: $\mathbf{x}^{[k]}(t) = \sum_{i=1}^n \mathbf{x}_i^{[k]}(t)$, where $\mathbf{x}_i^{[k]}(t) = \mathbf{A}_i^{[k]}\mathbf{s}_i^{[k]}(t)$. In a blind context, the component vector $\mathbf{x}_i^{[k]}(t)$ is better defined than the source vector $\mathbf{s}_i^{[k]}(t)$. Indeed, for any arbitrary invertible $m_i \times m_i$ matrix $\mathbf{Z}_i^{[k]}$, it is impossible to discriminate between the representation of a component $\mathbf{x}_i^{[k]}(t)$ by the pair $(\mathbf{A}_i^{[k]}, \mathbf{s}_i^{[k]}(t))$ and $(\mathbf{A}_i^{[k]}\mathbf{Z}_i^{-[k]}, \mathbf{Z}_i^{[k]}\mathbf{s}_i^{[k]}(t))$, where $\mathbf{Z}_i^{-[k]}$ denotes $(\mathbf{Z}_i^{[k]})^{-1}$. This means that only the column space of $\mathbf{A}_i^{[k]}$, $\mathrm{span}(\mathbf{A}_i^{[k]})$, can be blindly identified. Therefore, JISA is in fact a subspace estimation problem.

In the following, we focus on JISA using second-order statistics (SOS). Assuming sample independence $\forall t \ne t'$ and collecting $\widetilde{\mathbf{s}}(t) = [\mathbf{s}_1^\dagger(t), \ldots, \mathbf{s}_n^\dagger(t)]^\dagger$, the above model implies that $\widetilde{\mathbf{S}} \triangleq E\{\widetilde{\mathbf{s}}(t)\widetilde{\mathbf{s}}^\dagger(t)\} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{nn} \end{bmatrix} \triangleq \mathrm{bdiag}\{\mathbf{S}_{11}, \ldots, \mathbf{S}_{nn}\}$ is a $Km \times Km$ block-diagonal matrix with block-pattern $K\mathbf{m} = [Km_1, \ldots, Km_n]$. $\mathbf{S}_{ij} = E\{\mathbf{s}_i(t)\mathbf{s}_j^\dagger(t)\}$ is $Km_i \times Km_j$ for $1 \le i, j \le n$.

As we shall see shortly, it is useful to represent the entire set of observations as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \qquad (2)$$

where $\mathbf{s}(t) = [\mathbf{s}^{[1]\dagger}(t), \ldots, \mathbf{s}^{[K]\dagger}(t)]^\dagger$ and $\mathbf{x}(t) = [\mathbf{x}^{[1]\dagger}(t), \ldots, \mathbf{x}^{[K]\dagger}(t)]^\dagger$ are $Km \times 1$ vectors and $\mathbf{A} = \mathrm{bdiag}\{\mathbf{A}^{[1]}, \ldots, \mathbf{A}^{[K]}\}$. Accordingly, we obtain $\mathbf{X} = \mathbf{A}\mathbf{S}\mathbf{A}^\dagger$ where $\mathbf{S} = E\{\mathbf{s}(t)\mathbf{s}^\dagger(t)\}$ and $\mathbf{X} = E\{\mathbf{x}(t)\mathbf{x}^\dagger(t)\}$. With these notations, $\widetilde{\mathbf{s}}(t) = \mathbf{\Phi}\mathbf{s}(t)$, where $\mathbf{\Phi}$ is the corresponding $Km \times Km$ permutation matrix, and $\widetilde{\mathbf{S}} = \mathbf{\Phi}\mathbf{S}\mathbf{\Phi}^\dagger$.

We now discuss conditions under which blind identification of the component subspaces is possible. As mentioned in Sec. 1, one of the strong properties of JBSS is that it can (jointly) resolve several BSS problems using SOS alone, without further assumptions. Comparing the number of degrees of freedom (i.e. model unknowns) vs. number of constraints imposed by the data, as in [6, Sec. V.A], it is immediate to see that *generically*, for $K > 1$ the problem is always well-posed and thus identifiability of the model is guaranteed regardless of $\mathbf{m}$. On the other extreme, the same calculation shows that imposing statistical independence of all pairs $(\mathbf{s}_i^{[k]}, \mathbf{s}_i^{[k' \ne k]})$ yields a model that is never blindly identifiable using SOS, without further assumptions. The latter is not surprising, since such a model amounts to K separate BSS problems. Further discussion of (non-) identifiability of JISA is beyond the scope of this paper. Identifiability of classical JBSS is discussed in [14].

Our goal is component separation. Consequently, we define the figure of merit, for one set of observations, as the (normalized) mean square error (MSE) in the estimation of $\mathbf{x}_i(t) = [\mathbf{x}_i^{[1]\dagger}(t), \ldots, \mathbf{x}_i^{[K]\dagger}(t)]^\dagger$,

$$\widehat{\mathrm{MSE}}_i = \frac{1}{\sigma_i^2}\frac{1}{T}\sum_{t=1}^T \|\widehat{\mathbf{x}}_i(t) - \mathbf{x}_i(t)\|^2\,, \qquad (3)$$

where $\|\cdot\|$ denotes the Frobenius norm and $\sigma_i^2 = E\{\|\mathbf{x}_i(t)\|^2\}$. It can be shown (proof omitted) that estimates of $\mathbf{x}_i(t)$ obtained from matrices that satisfy the stationary points of (6) achieve asymptotically ($T \to \infty$) the minimal mean square error (MMSE).

2.1. Likelihood

In the following, we consider a Gaussian model in which $\mathbf{s}_i(t) \sim \mathcal{N}(\mathbf{0}_{Km_i \times 1}, \mathbf{S}_{ii})$ are mutually independent samples $\forall t \ne t'$. Using the Kullback-Leibler-induced divergence (KLiD) $D(\mathbf{R}_1, \mathbf{R}_2) = \frac{1}{2}(\mathrm{tr}\{\mathbf{R}_1\mathbf{R}_2^{-1}\} - \log\det(\mathbf{R}_1\mathbf{R}_2^{-1}) - m)$ between any two $m \times m$ positive definite matrices $\mathbf{R}_1$ and $\mathbf{R}_2$, the (minus, normalized) log-likelihood for the model just described can be written as [6, 16]

$$\begin{aligned} -\tfrac{1}{T}\log p(\mathcal{X}; \mathcal{A}, \mathbf{S}) &= \phi(\mathcal{A}, \mathbf{S}) = D(\overline{\mathbf{X}}, \mathbf{X}) + \kappa \\ &= D(\overline{\mathbf{X}}, \mathbf{A}\mathbf{S}\mathbf{A}^\dagger) + \kappa = D(\mathbf{\Phi}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\mathbf{\Phi}^\dagger, \widetilde{\mathbf{S}}) + \kappa \end{aligned} \qquad (4)$$

where $\mathcal{A} = \{\mathbf{A}^{[k]}\}_{k=1}^{K}$ and $\overline{\mathbf{X}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^{\dagger}(t)$ is the empirical counterpart of $\mathbf{X}$. The term $\kappa = \frac{1}{2}(\log\det(2\pi\mathbf{X}) + Km)$ is irrelevant to the maximization of the likelihood since it depends only on the data and not on the parameters. The third equality in (4) follows from (2). The fourth equality follows from the definitions of $\mathbf{\Phi}$ and $D(\cdot)$.

2.2. Derivation of the Relative Variations

We now derive the RG of the log-likelihood (4), which is the core of our algorithm. Given the block-diagonal structure of $\widetilde{\mathbf{S}}$, the last step in (4) gives rise [16] [6, Appendix B] to an ML estimate of the nuisance parameter $\mathbf{S}$, $\widehat{\widetilde{\mathbf{S}}}^{\mathrm{ML}} = \mathrm{bdiag}_{K\mathbf{m}}\{\mathbf{\Phi}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\mathbf{\Phi}^{\dagger}\}$. We can now write $\max_{\mathbf{S}}\log p(\mathcal{X};\mathcal{A},\mathbf{S}) = -TC(\mathbf{A}) + \kappa$ where we define the *contrast function*

$$C(\mathbf{A}) = D(\mathbf{\Phi}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\mathbf{\Phi}^{\dagger}, \mathrm{bdiag}_{K\mathbf{m}}\{\mathbf{\Phi}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\mathbf{\Phi}^{\dagger}\})$$

Since $D(\mathbf{R}, \mathrm{bdiag}_{\mathbf{m}}\{\mathbf{R}\}) \geq 0$ with equality iff $\mathbf{R} = \mathrm{bdiag}_{\mathbf{m}}\{\mathbf{R}\}$, minimizing the contrast amounts to a block-diagonalization problem under the constraint that $\mathbf{A}$ is block-diagonal.

The next step is to estimate $\mathcal{A}$. For this purpose, we calculate the derivative of $\phi(\mathcal{A},\mathbf{S})$ w.r.t. each $\mathbf{A}^{[k]}$ separately, for fixed $\mathbf{S}$ and $\mathbf{A}^{[l\neq k]}$, as we now explain. Consider a relative variation $\mathbf{A}^{[k]} \to \mathbf{A}^{[k]}(\mathbf{I}+\boldsymbol{\delta}^{[k]})^{-1}$, where $\boldsymbol{\delta}^{[k]}$ is $m\times m$ and has arbitrary values but such that $\mathbf{I}+\boldsymbol{\delta}^{[k]}$ is invertible[1]. Then, the first-order variation of $\phi(\mathcal{A},\mathbf{S})$ can be expressed by the Taylor expansion

$$\phi(\{\{\mathbf{A}^{[l]}\}_{l\neq k}, \mathbf{A}^{[k]}(\mathbf{I}+\boldsymbol{\delta}^{[k]})^{-1}\}, \mathbf{S}) = \phi(\mathcal{A},\mathbf{S}) + \mathrm{tr}\{(\nabla\phi^{[k]}(\mathcal{A},\mathbf{S}))^{\dagger}\boldsymbol{\delta}^{[k]}\} + O(\|\boldsymbol{\delta}^{[k]}\|^2), \quad (5)$$

where $\nabla\phi^{[k]}(\mathcal{A},\mathbf{S})$ denotes the $m\times m$ RG of $\phi(\mathcal{A},\mathbf{S})$ w.r.t. $\mathbf{A}^{[k]}$. Similarly to the derivation for multidimensional BSS in [6, Sec. III.D], one obtains that $\nabla\phi^{[k]}(\mathcal{A},\mathbf{S}) = \mathbf{E}_k^{\dagger}\mathbf{S}^{-1}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\mathbf{E}_k - \mathbf{I}_m$, where $\mathbf{E}_k \in \mathbb{R}^{Km\times m}$ consists of the adjacent m columns of $\mathbf{I}_{Km}$ that correspond to the columns of $\mathbf{A}^{[k]}$ in $\mathbf{A}$. The K terms $\nabla\phi^{[k]}(\mathcal{A},\mathbf{S})$ can be collected into $\nabla\phi(\mathcal{A},\mathbf{S}) = \mathrm{bdiag}_{\mathbf{k}}\{\mathbf{S}^{-1}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\} - \mathbf{I}$. It can be shown that the first-order variation of $C(\mathbf{A})$ w.r.t $\mathbf{A}^{[k]}$, derived similarly to (5), obeys $\nabla C^{[k]}(\mathbf{A}) = \nabla\phi^{[k]}(\mathcal{A},\mathbf{S})\big|_{\mathbf{S}=\widehat{\mathbf{S}}^{\mathrm{ML}}}$. Given the above, we can now write

$$\nabla C(\mathbf{A}) = \nabla\phi(\mathcal{A},\mathbf{S})\big|_{\mathbf{S}=\widehat{\mathbf{S}}^{\mathrm{ML}}} = \mathrm{bdiag}_{\mathbf{k}}\{\mathbf{\Phi}^{\dagger}\,\mathrm{bdiag}_{K\mathbf{m}}^{-1}\{\mathbf{\Phi}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\mathbf{\Phi}^{\dagger}\}\mathbf{\Phi}\mathbf{A}^{-1}\overline{\mathbf{X}}\mathbf{A}^{-\dagger}\} - \mathbf{I} \quad (6)$$

where $\mathrm{bdiag}_{\mathbf{m}}^{-1}\{\cdot\}$ stands for $(\mathrm{bdiag}_{\mathbf{m}}\{\cdot\})^{-1}$. Matrices that satisfy $\nabla C(\mathbf{A}) = \mathbf{0}$ are ML estimates of $\mathcal{A}$. This is the basis for our RG algorithm, which is described in Sec. 3.

[1] It should be emphasized that $\mathbf{A}$ is block-diagonal *by definition* and thus there is absolutely no meaning to perturbing its off-block-diagonal entries. This is the bifurcation point from which the derivation of the RG takes a different path than that in [6].

3. A RELATIVE GRADIENT ALGORITHM

We describe the proposed iterative RG algorithms in pseudocode, in Algorithm 1 and 2. The RG algorithm works as follows: according to (5), if $\boldsymbol{\delta}^{[k]}$ is a matrix with small enough values to ensure the invertibility of $\mathbf{I}+\boldsymbol{\delta}^{[k]}$, and if $\mathbf{A}^{[k]}$ is changed into $\mathbf{A}^{[k]}(\mathbf{I}+\boldsymbol{\delta}^{[k]})^{-1}$, then ϕ changes by the amount $\mathrm{tr}\{\nabla\phi^{[k]\dagger}\boldsymbol{\delta}^{[k]}\} + O(\|\boldsymbol{\delta}^{[k]}\|^2)$. Given $\boldsymbol{\delta}^{[k]} = -\lambda\nabla\phi^{[k]}$ and $\lambda > 0$ a real scalar, the update rule (line 5 in Algorithm 2) changes ϕ into $\phi - \lambda\|\nabla\phi\|^2 + O(\|\nabla\phi\|^3)$. Hence, the decrease of ϕ is guaranteed for small enough λ and $\boldsymbol{\delta}^{[k]}$ such that the higher-order terms in (5) become negligible. The update rule is iterated until $\|\nabla\phi\| \leq$ threshold. The transformation matrix $\mathbf{T}$ reflects the relative change in $\mathbf{A}$ at each iteration. The choice of the step-size in a RG algorithm determines its convergence rate, in terms of the number of required iterations. For our numerical experiments, we chose to set λ by a variant of the backtracking line search, in which its two parameters (see Algorithm 2) are drawn anew from a uniform distribution at each iteration: $\alpha \sim \mathcal{U}[0.01, 0.3]$, $\beta \sim \mathcal{U}[0.1, 0.8]$. This variation showed better convergence than fixing these parameters at the beginning of the run.

Algorithm 1 An iterative algorithm for JISA

1: **function** JISA($\overline{\mathbf{X}}$, $\mathbf{\Phi}$, K, $\mathbf{m}$, threshold)
2: $\quad\mathbf{A} \leftarrow \mathbf{I}$, $\mathbf{R} \leftarrow \overline{\mathbf{X}}$ ▷ Init
3: $\quad$**while** $\|\nabla C\| >$ threshold **do** ▷ (6)
4: $\quad\quad$Evaluate $\mathbf{T}$ ▷ Algorithm 2
5: $\quad\quad\mathbf{R} \leftarrow \mathbf{T}^{-1}\mathbf{R}\mathbf{T}^{-\dagger}$
6: $\quad\quad\mathbf{A} \leftarrow \mathbf{A}\mathbf{T}$ ▷ For output only
7: $\quad\quad\mathbf{Y} \leftarrow \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{R}\mathbf{A}^{-\dagger}\mathbf{\Phi}^{\dagger}$
8: $\quad\quad\nabla C \leftarrow \mathrm{bdiag}_{\mathbf{k}}\{\mathbf{\Phi}^{\dagger}\,\mathrm{bdiag}_{K\mathbf{m}}^{-1}\{\mathbf{Y}\}\mathbf{Y}\mathbf{\Phi}\} - \mathbf{I}$ ▷ (6)
9: $\quad$**end while**
10: $\quad$**return** $\mathbf{A}$
11: **end function**

Algorithm 2 Relative gradient update step

1: $\lambda \leftarrow 1$, draw α, β ▷ Init
2: **while** $C(\mathbf{A}(\mathbf{I}-\lambda\nabla C)^{-1}) > C(\mathbf{A}) - \alpha\lambda\mathrm{tr}\{\|\nabla C\|^2\}$ **do**
3: $\quad\lambda \leftarrow \beta\lambda$
4: **end while**
5: $\mathbf{T} \leftarrow (\mathbf{I}-\lambda\nabla C)^{-1}$

4. NUMERICAL VALIDATION

In this section, we illustrate the convergence properties of the algorithm, as well as its component-separation capacities, in numerical experiments. The real positive definite matrices $\mathbf{S}_{ii}$ are drawn as $\mathbf{S}_{ii} = \mathrm{diag}^{-\frac{1}{2}}\{\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\dagger}\}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\dagger}\,\mathrm{diag}^{-\frac{1}{2}}\{\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\dagger}\}$, where $\mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\dagger}$ is the singular value decomposition (SVD) of a $Km_i \times Km_i$

matrix whose i.i.d. entries $\sim \mathcal{N}(0,1)$. The underlying sources are created by right-multiplying the transpose of the Cholesky factorization of $\widetilde{\mathbf{S}}$ with $Km \times T$ i.i.d. $\sim \mathcal{N}(0,1)$ numbers. The stopping threshold is set to 10^{-6}, $\mathbf{m} = [1,4,5]$, $K = 6$ and $T = 2 \cdot 10^4$. The mixing is realized as $\mathbf{A}^{[k]} = \mathbf{I} + \boldsymbol{\Upsilon}$, where the entries of $\boldsymbol{\Upsilon}$ are zero-mean Gaussian i.i.d. with standard deviation (std) $1/5$ and drawn anew for each k. In all the following numerical experiments, $\mathbf{A}$ is initialized with $\mathbf{I}$. Since the component estimates are invariant to block-diagonal scaling (Sec. 2), we are concerned only about permutation errors. Permutation that is not properly corrected causes significant errors in component reconstruction. In our experimental results, we verified that such errors did not occur. This is due to the safety margins that we took in choosing the above-mentioned simulation parameters, that usually assure the proper convergence of the algorithm (Sec. 3). Further discussion of permutation ambiguity in JISA is beyond the scope of this paper. For "backwards compatibility" we verified that when the input to the algorithm is $\mathbf{m} = \mathbf{1}_{m\times 1}$, our algorithm achieves a similar optimum as the JBSS SOS algorithm in [16], up to acceptable numerical variations.

4.1. Convergence

We now illustrate the convergence properties of the algorithm. We keep $\mathbf{S}$ and $\overline{\mathbf{S}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{s}(t)\mathbf{s}^{\dagger}(t)$ fixed, only $\mathbf{A}$ varies at each of the $MC = 94$ trials. Fig. 1 shows the value of the contrast function (up) and number of iterations (bottom) when the stopping criterion is achieved. At each trial, the algorithm is run twice on the same data, in two modes: in the first mode, the input parameter $\mathbf{m}$ (Algorithm 1 line 1) reflects the true model structure. In the second mode, the input parameter is set to $\mathbf{1}_{m\times 1}$, a vector of all ones. The latter implies that the algorithm is ignorant of the true multidimensional structure of the data and instead tries to fit it to a one-dimensional model. This corresponds to *applying*, in a first step, a classical JBSS model to the data, with $m = 10 = n$ independent sources at each mixture (the number of latent sources is the same in both cases). In a second step, we cluster the output into the correct n multidimensional components before evaluating the contrast function. We denote this two-step approach "mismodeling".

It is clear from Fig. 1 (up) that in the correct model case, the algorithm converges to the same optimum (up to small variations that are due to the fact that the threshold is finite) regardless of $\mathbf{A}$. This follows from the definition of the KLiD. It is interesting to note that even in adverse conditions, that is, mismodeling, the algorithm does manage to converge to a rather narrow range of values. This problem becomes more severe as the mismodeling departs from the true data model. In addition, we note in Fig. 1 (bottom) that the mismodeled case usually requires a larger number of iterations to converge. This, too, is expected, since the algorithm is trying to block-diagonalize $\widetilde{\mathbf{S}}$ into smaller blocks than is actually possible and thus doing unnecessary work.

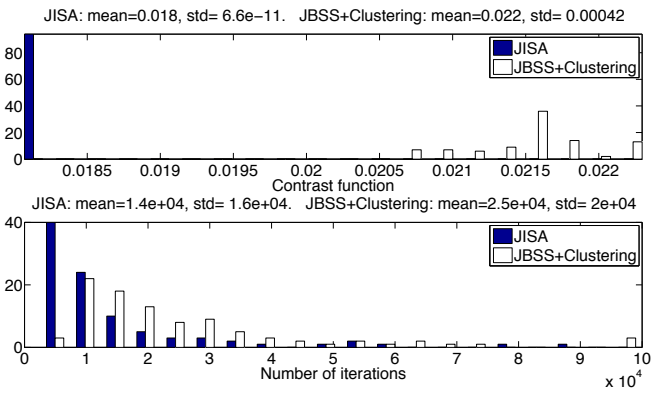

Fig. 1: Convergence of the RG algorithm. Histogram of the value of the contrast function (up) and number of iterations (bottom) when the stopping criterion is achieved.

4.2. Component Separation

We now evaluate the component separation quality of the proposed model and method. In the following experiment, we run multiple trials for fixed $\mathbf{S}$ (same as in Sec. 4.1) and $\mathbf{A}$ and varying $\overline{\mathbf{S}}$. The data is generated as explained above. As in Sec. 4.1, we compare JISA with JBSS. In the latter case, we cluster the output into the correct n multidimensional components before evaluating the component separation quality. For each trial we evaluate the normalized empirical MSE (3). Fig. 2 illustrates our results. Subplot i corresponds to component i. Each subplot shows the results of two different runs of the algorithm on the same data: once as true JISA with the correct data model, and once mismodeled, as explained above. For each of the two runs of the algorithm, the mean and std of $\widehat{\mathrm{MSE}}_i$ are written above the corresponding sublot. The histogram represents $MC = 100$ repetitions of this experiment. The small values of $\widehat{\mathrm{MSE}}_i$ in Fig. 2 confirm that the components have been properly separated. Fig. 2 shows that for all three components, there is a significant improvement (decrease) in MSE when the correct model is used w.r.t. the mismodeled scenario. This observation conforms with previous results on multidimensional components [6, 8].

5. DISCUSSION

This paper provides a "proof of concept" for a new data model that generalizes JBSS to multidimensional components. We demonstrate that JISA is capable of identifying real white Gaussian stationary multidimensional sources in multiple static mixtures, a task that is non-identifiable when only one such mixture is concerned or when there is no link between the mixtures. In JISA, this is made possible thanks to the additional diversity [14] offered by the statistical dependence between mixtures. We derive a RG non-orthogonal algorithm that is capable, as discussed in Sec. 2, of achieving asymptotically the MMSE in the joint estimation of sev-

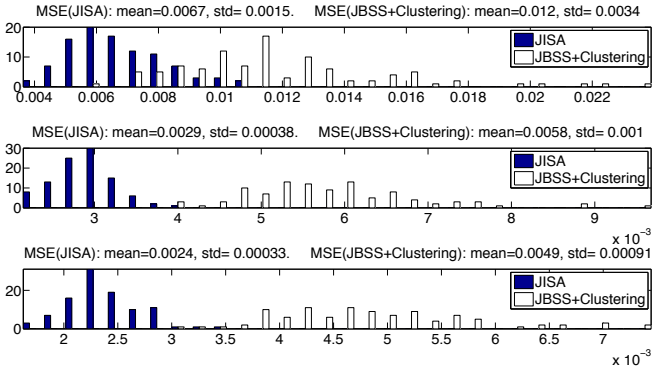

Fig. 2: Component separation. Histogram of the normalized empirical MSE. Subplots correspond to components with dimensions 1, 4 and 5, respectively.

eral Gaussian multidimensional components from their mixtures, based on minimizing a KLiD-based criterion. The algorithm can be applied to non-Gaussian data as well. However, in this case, the component estimation error will not achieve the MMSE. Simulations demonstrate the proper convergence of the algorithm when the small-errors regime is respected. In addition, we illustrate the gain in using JISA over classical JBSS followed by clustering, in terms of accuracy and number of iterations. Our analysis assumes that the dimension of the multidimensional components is known. At first sight this might seem as a drawback of our method. However, in fact, one has to keep in mind that although BSS [17] is a well-established method, and has proven useful in many applications [18, Chapters 16–19], the assumption that real-world signals can be modelled by one-dimensional sources is often just an *approximation*. Therefore, the fact that certain problems can be solved – *to some extent* – by classical BSS or JBSS, that is, an implicit rigid *choice* of dimension=1 to all the sources, does not mean that one should not be *allowed* to make other choices. Hence, the fact that multidimensional algorithms provide the user with the additional flexibility of choosing the data dimensions should not be regarded as a shortcoming of the algorithm but as means for a better fit to the data. The question of how to determine the correct dimension for the data, be it one or more, is beyond the scope of this work.

REFERENCES

[1] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Independent Component Analysis and Blind Signal Separation*, Heidelberg, 2006, vol. 3889 of *LNCS*, pp. 165–172, Springer.

[2] Y.-O. Li, T. Adalı, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3918–3929, Oct. 2009.

[3] J.-F. Cardoso, "Multidimensional independent component analysis," in *Proc. ICASSP*, Seattle, WA, May 1998, vol. 4, pp. 1941–1944.

[4] A. Hyvärinen and P. O. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Comput.*, vol. 12, no. 7, pp. 1705–1720, Jul. 2000.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by source subspace separation," in *Proc. IEEE SP/ATHOS Workshop on HOS*, Girona, Spain, Jun. 1995, pp. 134–138.

[6] D. Lahat, J.-F. Cardoso, and H. Messer, "Second-order multidimensional ICA: Performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4598–4610, Sep. 2012.

[7] D. Lahat, J.-F. Cardoso, and H. Messer, "Identifiability of second-order multidimensional ICA," in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 1875–1879.

[8] D. Lahat, J.-F. Cardoso, and H. Messer, "Blind separation of multidimensional components via subspace decomposition: Performance analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2894–2905, Jun. 2014.

[9] Z. Szabó, B. Póczos, and A. Lőrincz, "Separation theorem for independent subspace analysis and its consequences," *Pattern Recognition*, vol. 45, no. 4, pp. 1782–1791, Apr. 2012.

[10] S. Ma, N. M. Correa, Xi-Lin Li, T. Eichele, V. D. Calhoun, and T. Adalı, "Automatic identification of functional clusters in fMRI data using spatial dependence," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 12, pp. 3406–3417, Dec. 2011.

[11] J.-F. Cardoso, M. Le Jeune, J. Delabrouille, M. Betoule, and G. Patanchon, "Component separation with flexible models – application to multichannel astrophysical observations," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 5, pp. 735–746, Oct. 2008.

[12] X.-L. Li, T. Adalı, and M. Anderson, "Joint blind source separation by generalized joint diagonalization of cumulant matrices," *Signal Process.*, vol. 91, no. 10, pp. 2314–2322, Oct. 2011.

[13] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672–1683, Apr. 2012.

[14] T. Adalı, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Process. Mag.*, pp. 18–33, May 2014.

[15] S. Ma, V. D. Calhoun, R. Phlypo, and T. Adalı, "Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis," *NeuroImage*, vol. 90, pp. 196–206, 2014.

[16] J. Vía, M. Anderson, X.-L. Li, and T. Adalı, "A maximum likelihood approach for independent vector analysis of Gaussian data sets," in *Proc. MLSP*, Beijing, China, Sep. 2011.

[17] P. Comon, "Independent component analysis," in *Proc. Int. Signal Process. Workshop on HOS*, Chamrousse, France, Jul. 1991, pp. 111–120, Keynote address. Republished in *HOS*, J.-L. Lacoume ed., Elsevier, 1992, pp. 29–38.

[18] P. Comon and Ch. Jutten, Eds., *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, 1st edition, Feb. 2010.